

Kernel smoothing has greatly evolved since its inception to become an essential methodology in the Data Science tool kit for the 21st century. Its widespread adoption is due to its fundamental role for multivariate exploratory data analysis, as well as the crucial role it plays in composite solutions to complex data challenges.

Multivariate Kernel Smoothing and Its Applications offers a comprehensive overview of both aspects. It begins with a thorough exposition of the approaches to achieve the two basic goals of estimating probability density functions and their derivatives. The focus then turns to the applications of these approaches to more complex data analysis goals, many with a geometric/topological flavour, such as level set estimation, clustering (unsupervised learning), principal curves, and feature significance. Other topics, while not direct applications of density (derivative) estimation but sharing many commonalities with the previous settings, include classification (supervised learning), nearest neighbour estimation, and deconvolution for data observed with error.

For a Data Scientist, each chapter contains illustrative Open Data examples that are analysed by the most appropriate kernel smoothing method. The emphasis is always placed on an intuitive understanding of the data provided by the accompanying statistical visualisations. For a reader wishing to investigate further the details of their underlying statistical reasoning, a graduated exposition to a unified theoretical framework is provided. The algorithms for efficient software implementation are also discussed.

José E. Chacón is an associate professor at the Department of Mathematics of the Universidad de Extremadura in Spain.

Tarn Duong is a Senior Data Scientist for a start-up which provides short distance carpooling services in France.

Both authors have made important contributions to kernel smoothing research over the last couple of decades.



CRC Press
Taylor & Francis Group
an **informa** business
www.crcpress.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
711 Third Avenue
New York, NY 10017
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK



Preface	xiii
List of Figures	xv
List of Tables	xix
List of Algorithms	xxi
1 Introduction	1
1.1 Exploratory data analysis with density estimation	1
1.2 Exploratory data analysis with density derivatives estimation	4
1.3 Clustering/unsupervised learning	5
1.4 Classification/supervised learning	6
1.5 Suggestions on how to read this monograph	7
2 Density estimation	11
2.1 Histogram density estimation	11
2.2 Kernel density estimation	14
2.2.1 Probability contours as multivariate quantiles	16
2.2.2 Contour colour scales	19
2.3 Gains from unconstrained bandwidth matrices	19
2.4 Advice for practical bandwidth selection	23
2.5 Squared error analysis	26
2.6 Asymptotic squared error formulas	30
2.7 Optimal bandwidths	35
2.8 Convergence of density estimators	36
2.9 Further mathematical analysis of density estimators	37
2.9.1 Asymptotic expansion of the mean integrated squared error	37
2.9.2 Asymptotically optimal bandwidth	39
2.9.3 Vector versus vector half parametrisations	40

3	Bandwidth selectors for density estimation	43
3.1	Normal scale bandwidths	44
3.2	Maximal smoothing bandwidths	45
3.3	Normal mixture bandwidths	46
3.4	Unbiased cross validation bandwidths	46
3.5	Biased cross validation bandwidths	49
3.6	Plug-in bandwidths	49
3.7	Smoothed cross validation bandwidths	52
3.8	Empirical comparison of bandwidth selectors	54
3.9	Theoretical comparison of bandwidth selectors	60
3.10	Further mathematical analysis of bandwidth selectors	61
3.10.1	Relative convergence rates of bandwidth selectors	61
3.10.2	Optimal pilot bandwidth selectors	64
3.10.3	Convergence rates with data-based bandwidths	65
4	Modified density estimation	67
4.1	Variable bandwidth density estimators	67
4.1.1	Balloon density estimators	68
4.1.2	Sample point density estimators	69
4.1.3	Bandwidth selectors for variable kernel estimation	70
4.2	Transformation density estimators	73
4.3	Boundary kernel density estimators	76
4.3.1	Beta boundary kernels	76
4.3.2	Linear boundary kernels	77
4.4	Kernel choice	81
4.5	Higher order kernels	83
4.6	Further mathematical analysis of modified density estimators	84
4.6.1	Asymptotic error for sample point variable bandwidth estimators	84
4.6.2	Asymptotic error for linear boundary estimators	86
5	Density derivative estimation	89
5.1	Kernel density derivative estimators	89
5.1.1	Density gradient estimators	90
5.1.2	Density Hessian estimators	92
5.1.3	General density derivative estimators	93
5.2	Gains from unconstrained bandwidth matrices	96
5.3	Advice for practical bandwidth selection	100
5.4	Empirical comparison of bandwidths of different derivative orders	102

5.5	Squared error analysis	103
5.6	Bandwidth selection for density derivative estimators	108
5.6.1	Normal scale bandwidths	109
5.6.2	Normal mixture bandwidths	110
5.6.3	Unbiased cross validation bandwidths	111
5.6.4	Plug-in bandwidths	112
5.6.5	Smoothed cross validation bandwidths	115
5.7	Relative convergence rates of bandwidth selectors	117
5.8	Case study: The normal density	118
5.8.1	Exact MISE	118
5.8.2	Curvature matrix	119
5.8.3	Asymptotic MISE	120
5.8.4	Normal scale bandwidth	121
5.8.5	Asymptotic MSE for curvature estimation	122
5.9	Further mathematical analysis of density derivative estimators	124
5.9.1	Taylor expansions for vector-valued functions	124
5.9.2	Relationship between multivariate normal moments	124
6	Applications related to density and density derivative estimation	127
6.1	Level set estimation	127
6.1.1	Modal region and bump estimation	129
6.1.2	Density support estimation	132
6.2	Density-based clustering	135
6.2.1	Stable/unstable manifolds	136
6.2.2	Mean shift clustering	137
6.2.3	Choice of the normalising matrix in the mean shift	143
6.3	Density ridge estimation	143
6.4	Feature significance	149
7	Supplementary topics in data analysis	155
7.1	Density difference estimation and significance testing	155
7.2	Classification	159
7.3	Density estimation for data measured with error	163
7.3.1	Classical density deconvolution estimation	164
7.3.2	Weighted density deconvolution estimation	166
7.3.3	Manifold estimation	170
7.4	Nearest neighbour estimation	171
7.5	Further mathematical analysis	178

7.5.1	Squared error analysis for deconvolution kernel density estimators	178
7.5.2	Optimal selection of the number of nearest neighbours	179
8	Computational algorithms	181
8.1	R implementation	181
8.2	Approximate binned estimation	185
8.2.1	Approximate density estimation	185
8.2.2	Approximate density derivative and functional estimation	190
8.3	Recursive computation of the normal density derivatives	191
8.4	Recursive computation of the normal functionals	195
8.5	Numerical optimisation over matrix spaces	197
A	Notation	199
B	Matrix algebra	205
B.1	The Kronecker product	205
B.2	The vec operator	206
B.3	The commutation matrix	206
	Bibliography	207
	Index	225